

DT/KNN2 (Q10L11)

Total Questions: 20

Most Correct Answers: #9

Least Correct Answers: #14

1. This function can be used to perform KNN classification in R

- 6/7 ☒ A knn()
0/7 ☐ B k_nn()
1/7 ☐ C knnreg()
0/7 ☐ D knearneib()
0/7 ☐ E I do not know

2. With the increase of k, the decision boundary will be

- 2/7 ☒ A simplified
4/7 ☐ B more complex
0/7 ☐ C I do not know
1/7 ☐ D unchanged

3. In the case of small k we have

- 6/7 ☒ A overfitting
0/7 ☐ B underfitting
0/7 ☐ C it depends on the situation
1/7 ☐ D I do not know

4. Do you need to worry about scaling with one explanatory variable?

- 5/7 ☒ A No
2/7 ☐ B Yes
0/7 ☐ C I do not know

5. n - the number of observation,
 m - the number of explanatory variables

When $n=k$, $m=1$, the decision boundary for regression is

- 3/7 ☒ A a line
- 2/7 ☐ B a stepwise constant function
- 1/7 ☐ C a stepwise quadratic function
- 1/7 ☐ D I do not know

6. Which of these algorithms can be used to fill the missing values

- 2/7 ☐ A KNN for regression
- 3/7 ☐ B KNN for classification
- 1/7 ☒ C both
- 0/7 ☐ D I do not know

7. Which one is better: KNN regression or Linear regression
?

- 3/7 ☐ A KNN outperform LR if the parametric form that has been selected is close to the true form of f
- 4/7 ☒ B LR outperform KNN if the parametric form that has been selected is close to the true form of f
- 0/7 ☐ C KNN will always outperform the LR
- 0/7 ☐ D I do not know

8. Which one is the Disadvantage of KNN?

- 0/7 ☐ A required assumptions
- 0/7 ☐ B cannot be applied for regression
- 1/7 ☐ C difficult to perform
- 4/7 ☒ D the problem of high dimensional data
- 1/7 ☐ E I do not know

9. The best k for train set equals to

- 7/7 ☒ A 1
- 0/7 ☐ B 2
- 0/7 ☐ C 0
- 0/7 ☐ D I do not know

10. Decision tree is

- 4/7 ☒ A supervised learning algorithm
- 3/7 ☐ B unsupervised learning algorithm
- 0/7 ☐ C I do not know

11. Decision Tree Decision Boundaries

- 3/7 ☐ A are a step-wise constant function
- 2/7 ☐ B I do not know
- 1/7 ☐ C continuous function
- 1/7 ☒ D are axis-parallel rectangles

12. Root Node has

- 4/7 ☒ A no incoming edges and zero or more outgoing edges
- 0/7 ☐ B one incoming edge and two or more outgoing edges
- 0/7 ☐ C one incoming edge and no outgoing edges
- 3/7 ☐ D I do not know

13. Child or Internal Node has

- 0/7 ☐ A no incoming edges and zero or more outgoing edges
- 5/7 ☒ B one incoming edge and two or more outgoing edges
- 1/7 ☐ C one incoming edge and no outgoing edges
- 1/7 ☐ D I do not know

14. Pruning the tree means

- 1/7 ☒ A Simplify the tree
- 2/7 ☐ B Split the tree's nodes
- 1/7 ☐ C Merge the tree's nodes
- 3/7 ☐ D I do not know

15. Gini index equals to

- 1/7 ☒ A $1 - \sum (p_i^2)$
0/7 ☐ B $1 + \sum (p_i^2)$
2/7 ☐ C $\sum (p_i * \log(p_i))$
0/7 ☐ D $-\sum (p_i * \log(p_i))$
3/7 ☐ E I do not know

16. Entropy starts with 0

- 2/7 ☐ A True
1/7 ☒ B False
4/7 ☐ C I do not know

17. Overall impurity measure can be obtained by

- 3/7 ☒ A a weighted average of individual rectangles
2/7 ☐ B majority voting
2/7 ☐ C I do not know

18. At each stage, we choose the split with

- 3/7 ☒ A the lowest Gini index
0/7 ☐ B the lowest Chi-square value
0/7 ☐ C the highest entropy
4/7 ☐ D I do not know

19. We can perform the Decision Trees in r using

- 1/7 ☒ A `rpart()`
1/7 ☐ B `decisiontree()`
1/7 ☐ C `destree()`
3/7 ☐ D `reg.tree()`
1/7 ☐ E I do not know

20. minsplit in R means

- 2/7 ☒ A the minimum number of observations that must exist in a node in order for a split to be attempted
- 0/7 ☐ B the minimum number of observations in any terminal node
- 1/7 ☐ C the minimum number of splits
- 4/7 ☐ D I do not know